

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Факультет комп'ютерних наук та кібернетики

Кафедра теоретичної кібернетики

Дипломна робота

на здобуття ступеня бакалавра

за спеціальністю 122 Комп'ютерні науки

на тему:

**ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ АНАЛІЗУ ЕКГ ЗА
ДОПОМОГОЮ СТАТИСТИЧНОГО АНАЛІЗУ СИГНАЛІВ**

Виконав студент 4 курсу

Микитась Андрій Володимирович

(підпис)

Науковий керівник:

професор, доктор фіз.-мат. наук

Пашко Анатолій Олексійович

(підпис)

Засвідчую, що в цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент

(підпис)

Київ – 2021

РЕФЕРАТ

Обсяг роботи 42 сторінки, 13 ілюстрацій, 5 таблиць, 21 джерело посилань.

ЕЛЕКТРОКАРДІОГРАМА, ВАРІАБЕЛЬНІСТЬ СЕРЦЕВОГО РИТМУ, ЧАСТОТА СЕРЦЕВИХ СКОРОЧЕНЬ.

Об'єктом дослідження є аналіз та інтерпретація сигналів отриманих з електрокардіограм. Предметом дослідження є визначення поведінки серця за сигналом електрокардіограми.

Метою роботи є розробка системи для автоматизації обробки і аналізу електрокардіограм.

В роботі досліджені статистичні методи обробки даних та методи інтерпретації сигналів електрокардіограм. Були отримані наступні результати: виконано загальний огляд методів і інструментів статистичного аналізу електрокардіограм, проведена обробка даних ЕКГ та їх візуалізація. У цій роботі відома техніка, застосовується метод розкладання емпіричного режиму (ERP) для вилучення прихованої інформації в записаному сигналі ЕКГ. Тут намагаємось дослідити людське серце в динаміці, моделювати та виконувати ERP на звітах ЕКГ, процес пошуку таких захворювань, як аритмія, від нормальних отриманих даних із широко використовуваної бази даних МІТ-ВІН. ERP по суті включає розкладання сигналу в кінцеву кількість функцій внутрішнього режиму (ФВР), зберігаючи свої початкові властивості без змін. Для аналізу було використано фільтр Савіцького-Голея (SG) для видалення нестационарних шумів із сигналів ЕКГ. Популярний параметр Показник Херста (H) оцінюється для кожного ФВР методом R / S .

Застосовано ефективний підхід для обробки серцевого ритму на записах ЕКГ за двома категоріями: здоровий та аритмічний. Розроблено програмний засіб, що реалізує викладені у цій роботі алгоритми для обробки ЕКГ.

ЗМІСТ

ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ АНАЛІЗУ ЕКГ ЗА ДОПОМОГОЮ СТАТИСТИЧНОГО АНАЛІЗУ СИГНАЛІВ	1
ПЕРЕЛІК СКОРОЧЕНЬ ТА ВИЗНАЧЕНЬ	4
ВСТУП	5
РОЗДІЛ 1	8
ПІДГОТОВКА ДО АНАЛІЗУ ЕКГ	8
1.1. Загальні дані про аналіз сигналів та обробку ЕКГ	8
1.2. Збір даних ЕКГ	11
1.3. Згладження даних фільтром Савіцького – Голя.....	12
1.4. Адаптація ЕРР та вилучення ФВР	13
1.5. Визначення значних ФВР	16
1.6. Оцінка показників Херста	17
1.7. Кількісне дослідження показників Херста	19
1.8. Підгрупування набору даних на основі віку та статі	19
1.9. Розрахунок коефіцієнта кореляції - „момент продукту Пірсона (r)“....	19
РОЗДІЛ 2	22
ОБРОБКА ПРАКТИЧНИХ ДАНИХ	22
2.1 Перевірка роботи методів шляхом застосування на практичних даних	22
2.2 Розгляд результатів за різними статево-віковими групами.....	28
РОЗДІЛ 3	33
РЕАЛІЗАЦІЯ. ІНСТРУМЕНТИ ТА ЗАСОБИ.....	33
3.1 Інструменти та засоби.....	33
3.2 Реалізація мовою програмування Python.....	35
. ВИСНОВКИ.....	39
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	41

ПЕРЕЛІК СКОРОЧЕНЬ ТА ВИЗНАЧЕНЬ

ЕКГ - Електрокардіограма

ЕРР - Емпіричне розкладання режимів

ФВР - Функції власного режиму (IMF)

SG - фільтр Савіцького-Голя

H – Показник Херста

ВСТУП

Останнім часом спостерігається широке поширення захворювань серця у всьому світі. За даними Всесвітньої організації охорони здоров'я, серцево-судинні захворювання є однією з основних причин смерті у всьому світі. А звичайна та усталена методика оцінки стану серця включає використання сигналів електрокардіограми (ЕКГ). Інші сучасні методи, такі як тест Холтера, ехокардіограма та стрес-тест також використовуються для діагностики серцевого пацієнта. Однак наявність таких витончених і відносно дорогих засобів для виявлення серцевих захворювань обмежені здебільшого столичними та іншими великими містами країни. Незважаючи на цей факт, ЕКГ все ще відіграє важливу роль для кардіологів, щоб проаналізувати здоров'я серця простим та економічно вигідним способом. ЕКГ індивіда є графічним представлення електричного потенціалу серця, яке зазвичай використовується для виявлення наявності серцево-судинних розладів. Однак надати певну життєво важливу інформацію про серце складно лише за допомогою звичайного аналізу на основі ЕКГ за відсутності детальних знань щодо власних статистичних методів.. Таким чином обрана **тема дослідження є актуальною. Об'єктом дослідження є аналіз та інтерпретація сигналів отриманих з електрокардіограм, предметом дослідження є діагностування через аналіз ЕКГ.**

Метою даної роботи встановлення алгоритму, який допоможе виділити особливості, приховані в ЕКГ-сигналах для полегшення їх аналізу кардіологами без особливих знань щодо деталей математичних інструментів, пов'язаних з статистичними методами. Для досягнення цієї мети поставлено такі завдання.

1. Дослідити статистичні методи аналізу ЕКГ.
2. Виробити алгоритм обробки ЕКГ.
3. Перевірити ефективність роботи алгоритму на даних пацієнтів.
4. Розробити програмний засіб для автоматизації роботи алгоритму.

Часто використовуються такі методи, як мультифрактальний аналіз, вейвлет-перетворення, тощо. Однак, Емпіричне розкладання режимів (ЕРР), здається, є привабливим інструментом серед вищезазначених методів для ефективної боротьби з нелінійністю складної системи.

Варіації частоти серцевих скорочень зумовлені систолою та діастолою, індукованими спеціалізованими клітинами синоатріального вузла, атріовентрикулярного вузла разом із системою Гіса-Пуркінє . На основі частоти серцебиття, аритмія класифікується як (а) тахікардія (пульс у спокої перевищує 100 ударів на хвилини) та (б) брадикардія (частота серцевих скорочень у спокої нижче 60 ударів на хвилину), тоді як частота серцевих скорочень для здорової людини становить близько 72 ударів / хв.

До цих пір докладено зусиль для виявлення аритмії за допомогою різноманітного адаптивного фільтра конструкцій. Різні компоненти сигналу ЕКГ були враховані окремо, розглянуті та порівняні із звичайними сигналами ЕКГ для виявлення таких відхилень, як передчасні шлуночкові комплекси, фібриляція передсердь та інші.

Більше того, акцент також зроблено на класифікації аритмії на основі інтервалу RR, отриманого з сигналів ЕКГ. Діагноз цього захворювання 4 продовжує значною мірою залежати від інтуїтивного аналізу лікаря на основі звіти електрокардіограми (ЕКГ). Інтерпретація таких звітів може відрізнятись від лікаря до іншого і може призвести до неадекватного лікування з боку недосвідчених лікарів. Це має потенційні ризики для пацієнтів і може призвести до летальних наслідків. Більше того, час є обмеженням для інтерпретації звітів ЕКГ пацієнтів, які страждають від невідкладних медичних проблем, які вимагають аналізу цих звітів протягом декількох секунд. Більше того, для миттєвого моніторингу у невідкладних ситуаціях та інтерпретації ЕКГ пацієнта, час є життєво важливим фактором особливо коли рішення будуть прийматися за лічені секунди. Тому автоматична

інтерпретація медичних даних, таких як електрокардіограма, представляла інтерес дослідників за останні кілька десятиліть. Таким чином, шляхом просування в області обчислювальної біології, якість діагностики серцево-судинних захворювань за допомогою ЕКГ-сигналів можна було б покращити спільними зусиллями фізиків, математиків та медиків практики. У зв'язку з цим інструменти, що використовуються для діагностики, були б швидшими і з більшою ефективністю і меншим рівнем обслуговування, а головне, були б економічно вигідним. Протягом останніх двох десятиліть методи, засновані на статистиці, були застосовані для з'ясування динамічних особливостей, а саме невизначеності, системи інтересів. Серед різних типів даних медичні дані, такі як дані електрокардіограми (ЕКГ) та електроенцефалограми, дані про хворобу Паркінсона тощо виявляють значну кількість невизначеності. Дослідження в цій галузі включають методику, таку як біспектральний аналіз сигналів ЕКГ пацієнтів, які страждають фібриляцією передсердь виявленою на різних стадіях сну. Наше дослідження зосереджено на мінімізації вищезазначених труднощів шляхом розробки роману обчислювальні методи для виявлення пацієнтів з аритмією шляхом застосування методів ЕРР на ЕКГ-сигналах. З огляду на це, було зібрано дані ЕКГ пацієнтів, які страждають на аритмію і нормальні дані ЕКГ здорових людей з офіційного веб-сайту PhysioNet (PhysioNet). Застосування обчислювальної біології та наук про дані на різних біологічних комплексах системи дуже допомагають нам зрозуміти багато фізіологічних процесів. Різноманітні техніки такі як мультифрактальний аналіз, вейвлет-перетворення, тощо часто використовуються у фізіології та медичних науках. ЕРР здається привабливим інструментом серед вищезазначених методів. ЕРР по суті включає розкладання сигналу на кінцеву кількість ФВР, що зберігають його оригінальні властивості. Тут були визначені ФВР кожного ряду даних, (як у випадку захворювання, так і у протилежному)що обчислюється за допомогою застосування ЕРР і обчислюється показник Херста кожного ФВР.

РОЗДІЛ 1

ПІДГОТОВКА ДО АНАЛІЗУ ЕКГ

1.1. Загальні дані про аналіз сигналів та обробку ЕКГ

Електрокардіографія є одним з найрозповсюдженіших методів дослідження і діагностики серцево-судинної системи. Цей метод полягає у вимірюванні електричної активності серця з використанням електродів, розміщених на шкірі. Ці електроди виявляють малі електричні зміни, які є наслідком деполяризації серцевого м'яза з подальшою реполяризацією протягом кожного серцевого циклу (серцебиття). Зміна нормальної картини ЕКГ відбувається при численних порушеннях серцевої діяльності, включаючи порушення серцевого ритму (наприклад, фібриляція передсердь і шлуночкова тахікардія), недостатній кровотік коронарної артерії (наприклад, ішемія міокарда та інфаркт міокарда) і порушення електроліту (такі як гіпокаліємія і гіперкаліємія).). Таким чином ЕКГ дає можливість з великою точністю виявляти порушення роботи серця.

У звичайній ЕКГ десять електродів розміщують на кінцівках пацієнта і на поверхні грудей. Потім загальну величину електричного потенціалу серця вимірюють з дванадцяти різних кутів ("відведень") і записують протягом періоду часу (зазвичай десять секунд). Таким чином, загальна величина і напрямок електричної деполяризації серця фіксується в кожний момент протягом всього серцевого циклу. Є три головні компоненти ЕКГ: Р-хвиля, яка представляє деполяризацію передсердь; комплекс QRS, який представляє собою деполяризацію шлуночків; і Т-хвиля, що представляє собою реполяризацію шлуночків. Зазвичай на ЕКГ можна виділити 5 зубців: Р, Q, R, S, Т (рис. 1). Іноді можна побачити малопомітну хвилю U. Зубець Р відображає процес деполяризації міокарда передсердь, комплекс QRS - шлуночків, сегмент ST і зубець Т відображають процеси реполяризації міокарда шлуночків.

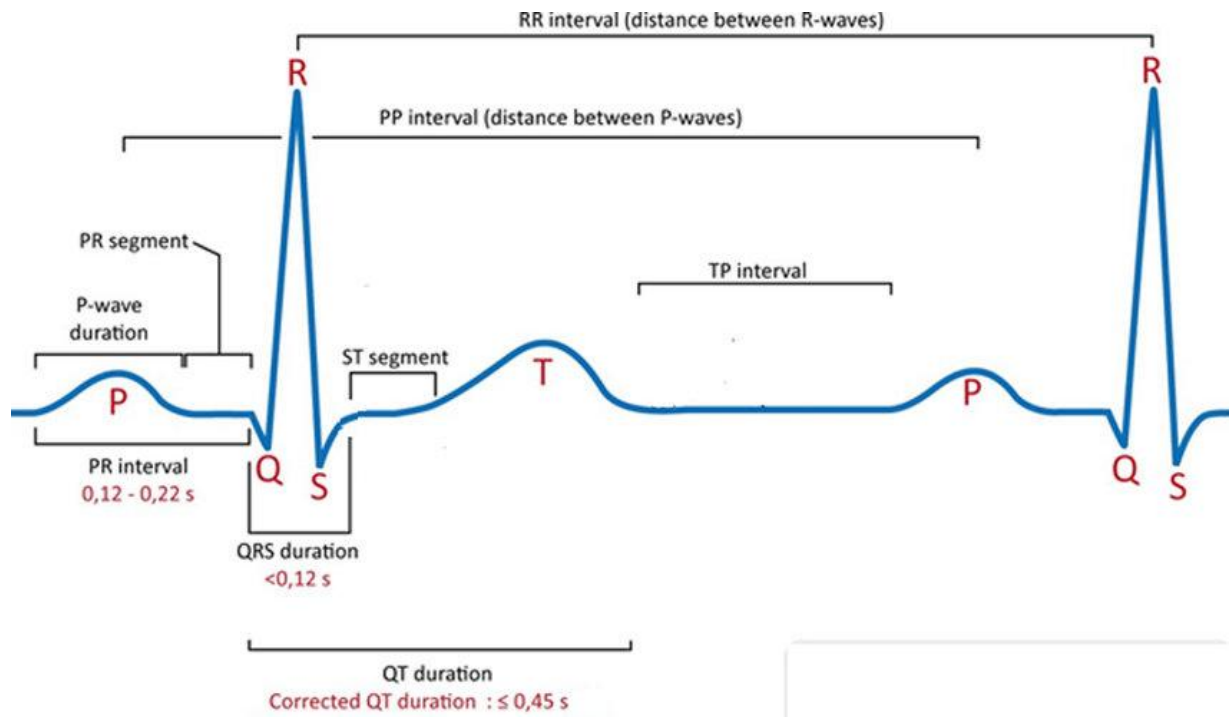


Рисунок 1 – структурні елементи кардіограми

Процес реполяризації - фаза, під час якої відновлюється вихідний потенціал спокою мембрани клітини після проходження через неї потенціалу дії. Під час проходження імпульсу відбувається тимчасова зміна молекулярної структури мембрани, в результаті якого іони можуть вільно проходити через неї. Під час реполяризації іони дифундують в зворотному напрямку для відновлення колишнього електричного заряду мембрани, після чого клітина виявляється готова до подальшої електричної активності.

Альтернативним методом є фотоплезитограма (ФПГ) (рис. 2). В цьому способі оптичні сенсори за допомогою світлодіода фіксують зміну об'єму крові внаслідок удару серця. ФПГ часто отримують за допомогою пульсоксиметра, який освітлює шкіру і вимірює зміни поглинання світла [2]. На сьогодні цей метод є популярним для використання у різного роду фітнес трекерах.

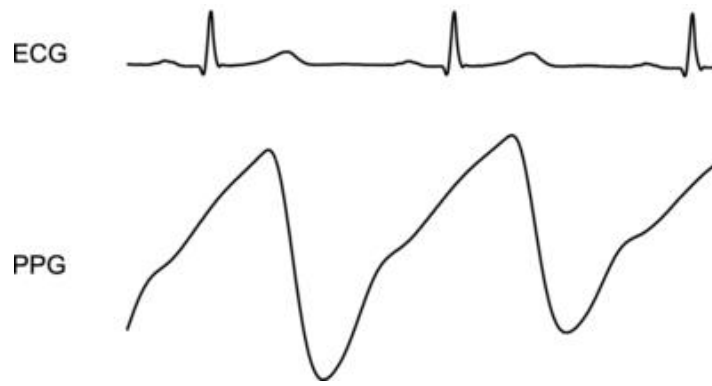


Рисунок 2

Серце людини - це складна система, що виявляє стохастичну природу, як це відображено в сигнали електрокардіограми (ЕКГ). ЕКГ-сигнал - це слабкий, нестационарний та нелінійний сигнал, що вказує на здоров'я серця з точки зору часових змін електромагнітних імпульсів від серця. Аномальні коливання сигналу ЕКГ викликають можливість різних серцево-судинних розладів, що діагностується за допомогою інтуїтивного аналізу звітів про ЕКГ лікарями. Це можна зробити швидко, точно і просто, наклавши вдосконалені статистичні інструменти на записаних ЕКГ-сигналах. У цій роботі відома техніка, застосовується метод розкладання емпіричного режиму (ЕРР) для вилучення прихованої інформації в записаному сигналі ЕКГ. Тут намагаємось дослідити людське серце в динаміці, моделювати та виконувати ЕРР на звітах ЕКГ, відрізняючи аритмію від нормальних отриманих даних із широко використовуваної бази даних МІТ-ВІН. ЕРР по суті включає розкладання сигналу в кінцеву кількість функцій внутрішнього режиму (ФВР), зберігаючи свої початкові властивості без змін. Для аналізу було використано потужний фільтр Савіцькі-Голея (SG) для видалення нестационарних шумів із сигналів ЕКГ. Популярний параметр Показник Херста (H) оцінюється для кожного ФВР методом R / S . Було визначено чітке поле H 1-го ФВР між нормою та аритмією страждає від пацієнтів. Ця модель підтверджує за допомогою 94,92% достовірність шансів виникнення хвороби аритмії у пацієнтів шляхом діагностики ЕКГ-сигнали без виконання інших

дорогих та трудомістких методів, таких як тест Холтера, ехокардіограма та стрес-тест.

1.2. Збір даних ЕКГ

З метою розрізнення людини, яка страждає аритмією, від здорової (нормальних) індивідуальних даних ЕКГ у повному обсязі для 48 аритмічних пацієнтів зібрано з популярної бази даних аритмії МІТ-ВІН (mitdb) (PhysioNet / МІТВІН_AD / 2018). Набори даних включають дані часових рядів ЕКГ 24 чоловіків у віці від 32 до 89 років та 22 жінок у віці від 23 до 89 років (Moody and Mark, 2001; Goldberger et співавт., 2000). З іншого боку, нормальні дані (здорова людина) були зібрані з бази даних нормального синусового ритму МІТВІН (nsrdb) (PhysioNet / МІТ-ВІН_NS RD / 2018), що складається з 18 тривалих ЕКГ-сигналів суб'єктів, які не мають значної аритмії. Предмети включають 5 чоловіків та 13 жінок у віці від 26 до 45 та від 20 до 50 відповідно (Goldberger et al., 2000). Сигналом, що використовується тут для аналізу, є модифікований відвід кінцівки II (MLII), отриманий розміщення електродів на грудях пацієнтів. Ми не включили записи двох пацієнтів з ідентифікаторами пацієнтів 102 та 104 з бази даних аритмії МІТ-ВІН як необхідні дані MLII не були доступні через хірургічні втручання у вищезазначених пацієнтів. З іншого боку також розглянемо дані з режиму ЕКГ1, які є сигналами ЕКГ (Нормальний синус База даних ритму), що стосується здорових людей, і ці набори даних розглядаються як доповнює дані MLII бази даних аритмії. Тут загальний номер точок даних кожного серії даних захворювань становить 21600 з частотою 360,01 сек-1, тоді як те ж

саме для норми ряд даних - 7680 з частотою 128 с^{-1} . Тому кожен із рядів даних записується для Тривалість часу 60 сек. Більше того, варіація сигналів ЕКГ у аритмічних пацієнтів є обмежена між $-3,105$ і $2,8$ мВ, тоді як вона коливається від $-2,215$ до $3,675$ мВ у випадку нормальних пацієнтів.

1.3. Згладження даних фільтром Савіцького – Голея

Фільтр використовується для обробки сигналів з метою вибіркової ізоляції конкретної частоти або діапазона частот з асортименту декількох частот у сигналі. Вибір відповідного фільтра для обробки системи сформованих сигналів вимагає максимального зменшення шуму з мінімальним спотворенням сигналу. Таким чином, проводяться масштабні дослідження щодо вибору цифрового фільтра з метою зменшення шуму, також це допомагає значно покращити результати. Для цього використовується кілька методів фільтрування шуму під час попередньої обробки біомедичних даних у режимі реального часу. Один з найпростіших фільтрів для згладжування часових рядів - це ковзне середнє. Загалом в середньому береться фіксована кількість точок, що називається фреймом, і додаються їх ординати з подальшим діленням на кількість балів, щоб отримати середню ординату. Потім, точка в одному кінець кадру скидається, а додається наступна крапка на іншому кінці. Цей процес виконується неодноразово до досягнення кінцевої точки часового ряду. Але, цей метод іноді буває явно небажаним через супровід деградації пікових інтенсивностей (Савіцький і Голей, 1964). Таким чином, один з найкращих фільтрів для очищення від шуму біомедичних даних, в т.ч. ЕКГ-сигналів, здається, є фільтр Савіцького – Голея (SG) (Харгіттай, 2005; Савіцькі та Голей, 1964). Основоположним принципом SG фільтра є врахування $2n + 1$ рівновіддалених точок $(-n, \dots, 0, \dots, + N)$ беручи $n = 0$ як центр для представлення полінома ступеня p (де p менше $2n + 1$). набір точок повинен бути підігнаний до деякої кривої. Для цього фільтр SG обчислює значення найменшого квадратного многочлена (або його

похідної) у точці, $i = 0$, у визначеному діапазоні кадру. Цей фільтр застосовує метод лінійних найменших квадратів для згладжування даних, що допомагає підтримувати вихідну форму сигналу (Sahoo et al., 2017) (рис. 3). Хочеться звернути увагу, що збільшення порядку та зменшення розміру кадру SG фільтра призводить до кращого збереження коротких коливань при одночасному непотрібному використанні полінома високого порядку з надзвичайно коротким розміром кадру може призвести до надмірного припасування. Загалом фільтр SG вимагає заздалегідь визначених значень порядку та кадру залежно від частоти та довжини даних. Зазвичай для вирішення цього питання потрібен метод проб і помилок або попередній досвід для визначення задовільних значень параметрів. Значення таких параметрів розглянуто в нашому дослідженні та інша необхідна інформація наборів даних також включена, як показано в таблиці 1.

Choice of order and frame of SG filter	Disease	Normal
Time period (sec)	0.0027777	0.0078125
Length of data	21600	7680
Frame of SG filter	37	13
Order of SG filter	3	3

Таблиця 1 - Значення різних параметрів, що враховуються для фільтру SG даних часових рядів

1.4. Адаптація EPP та вилучення ФВР

EPP - метод, розроблений Н. Хуангом, який зазвичай використовується для аналізу явищ що мають нелінійні та нестационарні характеристики. Цей метод передбачає розкладання вихідного сигналу на кінцеву кількість компонентів, повільно змінюючи амплітуду і фазу, що називається функцією внутрішнього режиму (ФВР). ФВР представляє коливальну частину сигналу. EPP є адаптивним методом завдяки виробництву сигнальних ФВР, які візуально дуже близькі до вихідних компонентів сигналу, з абсолютно

нерозрізнюваними частотами та амплітудами. Це багатомасштабний широко використовуваний метод аналізу, який широко використовується для прогнозування тенденції сигналу, не покладається будь-які попередні знання (Blanco-Velasco et al., 2008; Tolwinski, 2007). Це керовано даними механізмами, тоді як інші відомі методи статистичного аналізу даних, такі як Фур'є та вейвлет-методи, потребують деяких заздалегідь визначених базових функцій для подання сигналу. Процес отримання ФВР від сигналу називається процесом просіювання. Весь процес описаний наступним чином:

1. $x(t)$ розглядається як сигнал, а ініціювання процесу просіювання починається з ідентифікації всіх локальних максимумів та мінімумів $x(t)$.
2. Тоді всі локальні максимуми слід з'єднати кубічною сплайн-кривою з використанням інтерполяції, названий верхньою оболонкою $e_u(t)$. Подібним чином локальні мінімуми з'єднуємо, щоб отримати нижню оболонку $e_l(t)$.
3. Середнє значення верхньої та нижньої оболонки обчислюється та позначається $m_1(t)$:

$$m_1(t) = \frac{[e_u(t) + e_l(t)]}{2}$$

4. Потім середнє значення віднімається від основного сигналу $x(t)$, щоб отримати перший прото-ФВР $h_1(t)$. Таким чином,

$$h_1(t) = x(t) - m_1(t)$$

5. Через безліч екстремумів, присутніх між двома послідовними нульовими переходами, просіювання процес повинен застосовуватися безперервно до $h_k(t)$, прото-ФВР під номером k . Якщо це задовольняє Умови ФВР, перший ФВР $c_1(t)$ отримуємо
6. Процес просіювання зупиняється після досягнення критерію зупинки, який є характеризується сумою відхилень (SD) як

$$SD = \sum_{t=0}^T \frac{|h_{k-1}(t) - h_k(t)|^2}{h_{k-1}^2(t)}$$

Перший ФВР $c_1(t)$ отримується, коли SD менше або дорівнює пороговому значенню SD_{\max} . Типові значення SD_{\max} лежать в межах від 0,2 до 0,3.

7. Після цього ФВР виводиться з вихідного сигналу, щоб знайти перший залишковий сигнал $r_1(t)$ за допомогою

$$r_1(t) = x_1(t) - c_1(t)$$

8. Цей залишковий сигнал розглядається як вихідний сигнал для подальшого формування пари ФВР та залишкові сигнали. Процедуру можна продовжувати до залишку під номером N $r_N(t)$ перетворюється на константу або з одним екстремумом або з монотонним нахилом.

9. Нарешті, поєднуючи всі вищезазначені кроки, вихідний сигнал може бути виражений як

$$x(t) = \sum_{n=1}^N c_n(t) + r_n(t)$$

Кожен ФВР повинен задовольняти наступним характеристикам, як описано нижче:

1. Кількість перетину нуля повинна бути рівною або відрізнятись на одну одиницю від кількості екстремумів (припускаючи, що він має принаймні два екстремуми).
2. Кожен ФВР повинен бути симетричним щодо місцевого середнього значення.

1.5. Визначення значних ФВР

Отримавши всі ФВР, слід вибрати найважливіший від них. Для цього кореляція (C_{xcn}) між ФВР та відповідні вихідні дані часових рядів обчислюються з використанням наступного рівняння

$$C_{xcn} = \frac{\int x(t)c_n(t)}{\sqrt{\int x^2(t)dt \int c_n^2(t)dt}}$$

Жорсткий поріг λ використовується для вибору значущих ФВР. λ - відношення максимального C_{xcp} описано як

$$\lambda = \frac{\max(C_{xcp})}{\eta}$$

Де η - коефіцієнт відношення, що перевищує 1,0. У нашому дослідженні розглядається $\eta = 25$ відповідно до нашої зручності після методу спроб і помилок. ФВР із $C_{xcp} \geq \lambda$ були вважається значним. Для розрахунку кореляції використовували MATLAB R2016a коефіцієнти.

1.6. Оцінка показників Херста

Показник Херста (H) є одним із важливих факторів для перерахування міри нерегулярності, що існує в часовому ряді. Цей показник ввів популярний гідролог Е. Херст у 1951 р. Під час вирішення проблем управління водосховищами дамб на річці Ніл. Пізніше Мандельброт і Уолліс пояснили цю методологію щоб перевершити автокореляцію, ANOVA, а також спектральний аналіз (Mandelbrot і Уолліс, 1969). Чи значення H більше або менше 0,5, вказує на характер незв'язності набору даних. Коли значення H дорівнює 0,5, це називається випадковою прогулянкою.

Потім процес слідує за броунівським рухом. Якщо значення показника показника більше 0,5, це свідчить про стійку закономірність, тобто часові ряди мають довгострокову позитивну кореляцію, і вона буде регулярніше зберігати свої тенденції. На навпаки, значення H менше 0,5 вказує на стійкий характер. На кінцях, коли $H \rightarrow 0$, дані, як правило, не відповідають шаблону і змінюються непередбачувано, нерегулярно, демонструючи особливості білого шуму. З іншого боку, $H = 1$ відноситься до передбачуваного характеру набору даних. Коли функціональний контраст між окремими ФВР неможливо окреслити за допомогою основних законів статистики та закономірностей у зашумлених високочастотних даних, H може бути

потужним показником для характеристика пацієнтів. У цій роботі буде представлено досить новий підхід використання H для вилучення ознак усіх ФВР, отриманих із звітів ЕКГ аритмічних пацієнтів, а також здорових людей. У часових рядах доступні кілька методів аналізу, щоб з'ясувати H з даних часових рядів Один з найкращих методів оцінка H - R / S техніка, запропонована Мандельбротом і Уоллісом в 1969 р. Показник Херста, параметр самоподібності, забезпечує потужність довготривалої залежності в ФВР.

Етапи отримання H з допомогою методу R / S описані нижче:

1. Аналіз починається з розподілу ФВР довжини L на d підсерії ($z_{i,m}$) довжини n .
2. Середнє значення вибірки (E_m) та стандартне відхилення (S_m) розраховуються для кожної підсерії, $m = 1, \dots, d$
3. Тоді дані підсерії ($z_{i,m}$) нормуються як:

$$X_{i,m} = z_{i,m} - E_m \text{ for } i = 1, \dots, n.$$

4. Сукупний часовий ряд створюється як:

$$Y_{i,m} = \sum_{j=1}^i X_{j,m} \text{ for } i = 1, \dots, n.$$

5. Діапазон обчислюється із сукупного ряду як:

$$R_m = \max \{ Y_{i,m}, \dots, Y_{n,m} \} - \min \{ Y_{1,m}, \dots, Y_{n,m} \}.$$

6. Потім діапазон повторно масштабується шляхом ділення на S_m .
7. Нарешті, для всіх підсерій довжини n середнє значення перемасштабованого діапазону вважається таким:

$$\left(\frac{R}{S}\right)_n = \frac{1}{d} \sum_{m=1}^M \frac{R_m}{S_m}$$

8. Статистика R / S асимптотично слідує відношенню $(R / S)_n \approx cn^H$.

Значення H може бути оцінено простою лінійною регресією за допомогою рівняння:

$$\log(R/S)_n = \log c + H \log n$$

1.7. Кількісне дослідження показників Херста

Для пояснення результату експерименту були розраховані середнє та медіана H ФВР. Було проведено двосторонній t-тест студента з неоднаковими дисперсіями ($p < 0,05$), щоб ідентифікувати суттєву відмінність H 1-го ФВР аритмічних хворих від нормальних. Значення p відноситься до ймовірності будь-якої події, що сталася випадково внаслідок експериментальної несправності. Значення 0,05 або нижче вказує на значущі дані

1.8. Підгрупування набору даних на основі віку та статі

Набори даних були згруповані у дві підгрупи на основі віку та статі. За віком Група складається з чотирьох підгруп із до 30, 30-50, 50-70 та старше 70. Хоча гендерні групи були розділені для чоловіків та жінок. Вищезазначені кількісні процедури (див. 2.6) також дотримувались для кожної підгрупи.

1.9. Розрахунок коефіцієнта кореляції - „момент продукту Пірсона (r)“

У цій роботі зроблена спроба окреслити ступінь асоціації між віком суб'єктів (як хворих, так і нормальних) та відповідні значення H 1-го ФВР за допомогою статистичних підходів. Для цілей аналізу два різних записаних набору даних, тобто дві змінні, а саме: (а) вік хворих та здорових осіб та (б)

відповідний N були враховані значення 1-го ФВР. Коефіцієнти кореляції обчислюються за засобами оцінки найпоширенішого добре знайомого коефіцієнта кореляції: "момент продукту Пірсона" що визначається наступним чином):

$$r = \text{cov}(x, y) / (\sigma(x) \sigma(y)),$$

$$\text{cov}(x, y) = \sum (x_i - \bar{x}) (y_i - \bar{y}) / (N - 1)$$

де, $i = 1, 2, \dots, N$ і N довжина набору даних;

x_i (or y_i) = значення номер i від x (or y).

\bar{x} і \bar{y} , є середнім значенням та $\sigma(x)$, $\sigma(y)$ - стандартні відхилення наборів даних x і y , відповідно

Важливі статистичні інструменти, які часто використовуються для розшифровки лінійного взаємозв'язок між двома змінними включає коефіцієнт кореляції (r). За визначенням кажуть, що змінні є як лінійними, так і позитивними або корельованими певною мірою позитивно, якщо значення ' r ' дорівнює $+1$ або будь-якій додатній частці, тобто великі значення x пов'язані з вищими значеннями y і навпаки. З іншого боку, якщо значення ' r ' дорівнює -1 або будь-якому від'ємному числу, кажуть, що ці дві змінні є як лінійними, так і негативними або певною мірою корельованими негативно, тобто високі значення x пов'язані з низькими значеннями y і навпаки. $r = 0$, коли розглядається дві змінні, які не мають кореляції, тобто змінні мають випадковий характер. Більше того, крім математичного підходу, геометричне відношення між двома змінними x та y можуть бути графічно розроблені за допомогою графіків розсіювання разом з еліпсом довіри. Загалом, форма впевненого еліпса з розсіяним сюжетом посилається

на можливу природу кореляції. Наприклад, випадкові дані мають набагато більшу тенденцію до кругової форми, а позитивно / або негативно корельовані дані представлені вузькою формою еліпса. Значення коефіцієнта кореляції (r) визначається нахилом головної осі еліпса. Позитивні, негативні та нульові значення нахилу вказують на позитивний, негативний характер кореляції та некорельовані дані відповідно (Chaudhuri et al., 2019). Ціла методологія супроводжується принциповою схемою, як показано на рис.3.

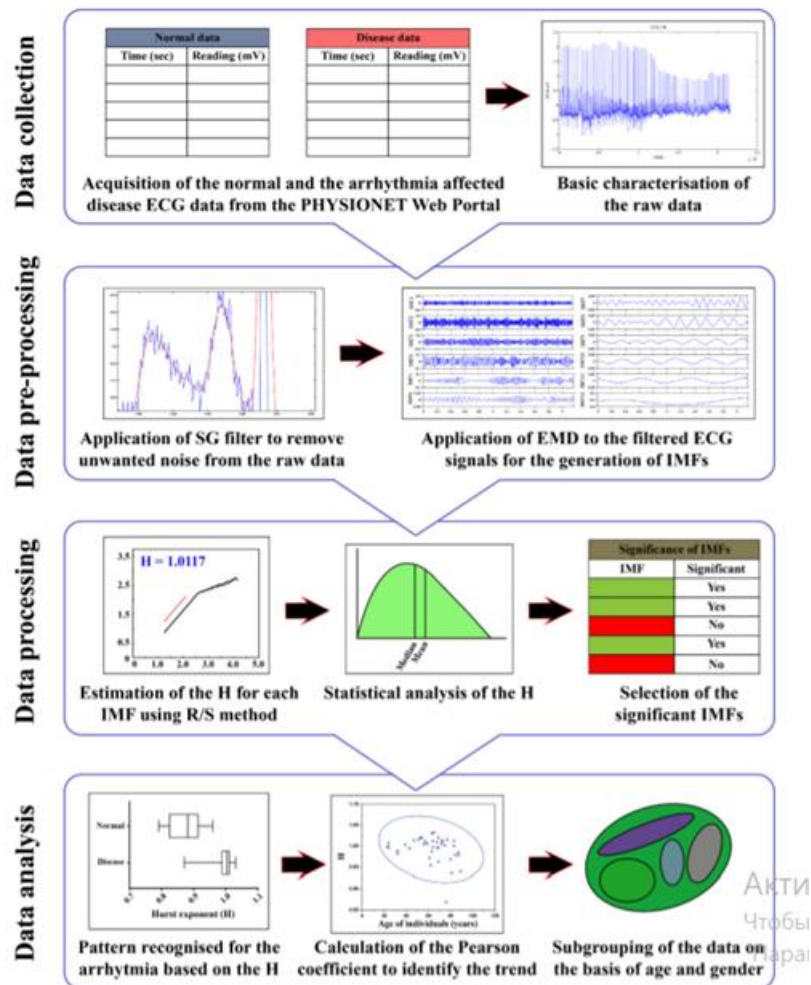


Рисунок 3 - Блок-схема методології, застосована у дослідженні

РОЗДІЛ 2

ОБРОБКА ПРАКТИЧНИХ ДАНИХ

2.1 Перевірка роботи методів шляхом застосування на практичних даних

Фільтр SG був застосований до даних часового ряду 46 пацієнтів з аритмією (так звана «хвороба».) І 18 нормальних пацієнтів для усунення шуму, присутнього в наборах даних. Тимчасовий варіація згладжених наборів даних одного здорового пацієнта (ID пацієнта: 19090) і одного хворого пацієнта (ID пацієнта: 100) нанесені на графік, як показано на рис. 4. Малюнок відображає відмінність між електрокардіограмою нормального і хворого пацієнта. Тут нормальна ЕКГ містить 47 піків за 60 секунд, тоді як 74 піка присутні в даних про хворого в такий же інтервал. Крім того, кожен цикл (довжина хвили) повторюється аналогічно на нормальній ЕКГ. 13 наборів даних є в наявності, але цього не спостерігається в наборах даних про хвороби. Значення максимуму (і мінімуму) амплітуди всіх циклів майже постійна протягом всього часу вимірювання для нормальних даних. Однак на ЕКГ даних хвороби спостерігається швидка зміна того ж самого. Отже, з первинного спостереження на графіку ЕКГ на рис. 2 можна зрозуміти відхилення ЕКГ хворого пацієнта порівняно з нормальним. Однак це не дає кількісної оцінки серйозності та типу захворювання. Тому аналіз ЕРР застосовували до кожного з них ряди даних (всього 64 серії, додаткова таблиця:) після процесу просіювання який описувався раніше для отримання ФВР. Тут максимальна кількість використаних ітерацій становить 150, а також отримали 14 - максимальну кількість ФВР. Кількість ФВР, отриманих з усіх серій, різняться між 10 і 14. ФВР також формуються як новий часовий ряд із рівними точками даних до оригінального часового ряду. Більше того, часові ряди ФВР відповідають одиниці Даних про хвороби (ідентифікатор пацієнта: 100) побудовані, як показано на рис. 5. Тепер, Н усіх ФВР усіх часові ряди

(Таблиця 2) були оцінені (рис. б) для перерахування (а) регулярності, (b) масштабування в рядах даних

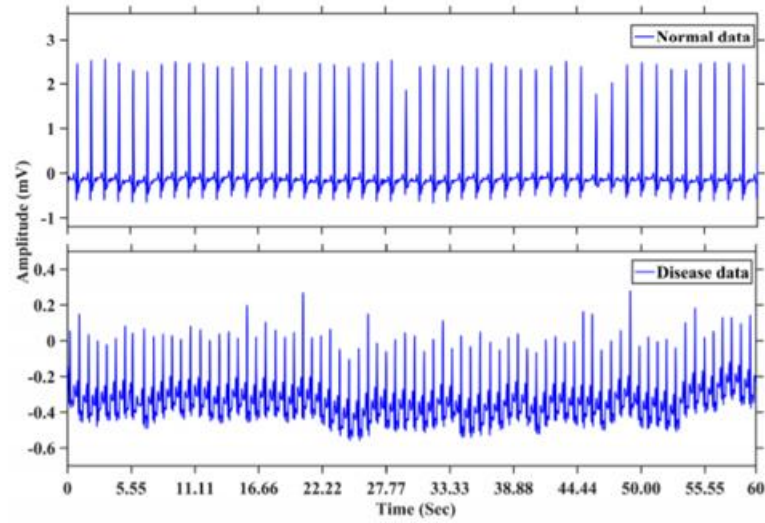


Рисунок 4 - Часова зміна даних ЕКГ одного здорового (Ідентифікатор пацієнта: 19090) та одного хворого

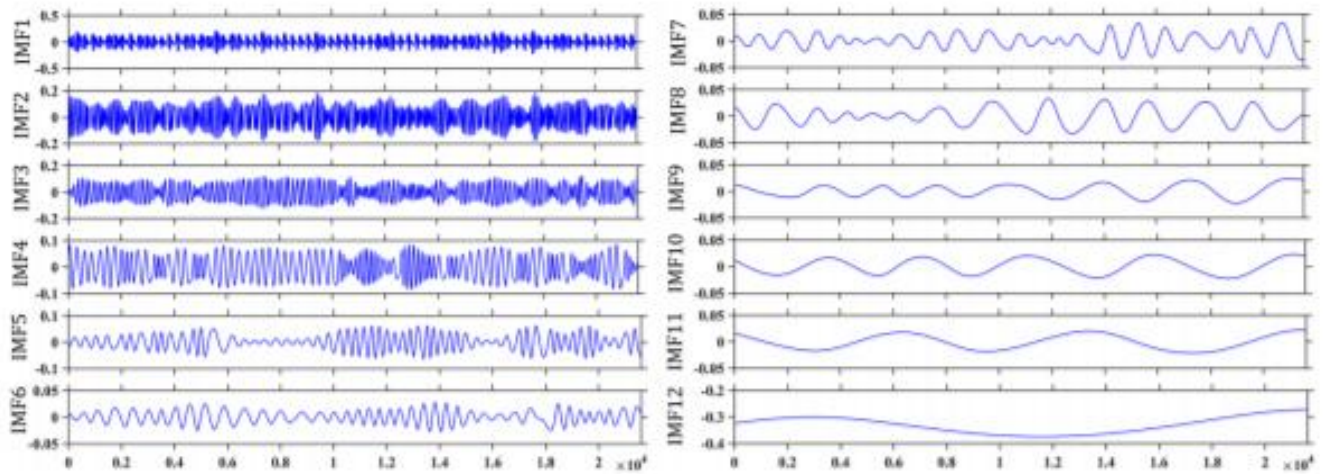


Рисунок 5 - Графіки всіх ФВР, що відповідають даним одного хворого пацієнта

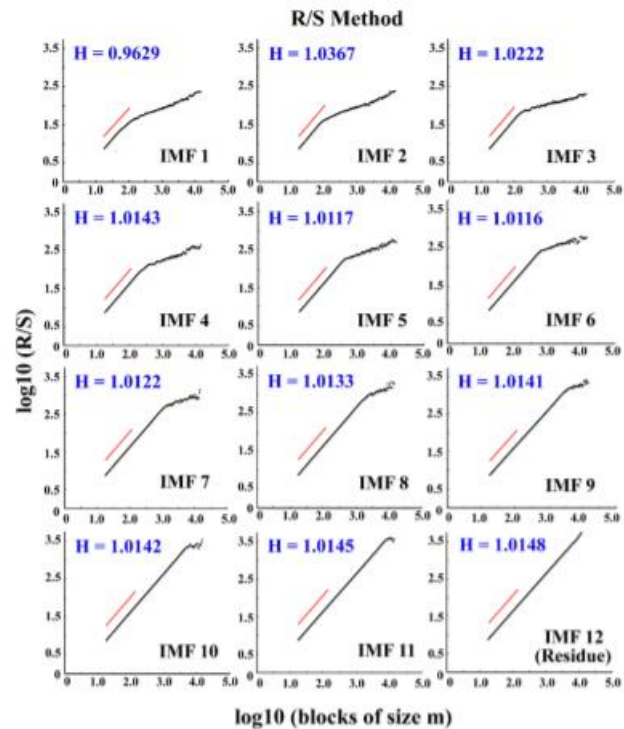


Рисунок 6 - Оцінка H усіх ФВР даних хворого пацієнта (Ідентифікатор пацієнта: 100) за даними R / S техніки

Порівняння H різних ФВР для даних нормальної (Ідентифікатор пацієнта: 19090) та хвороби (Ідентифікатор пацієнта: 100) наведено в Таблиці 2. Як підпис аритмії щодо H спостерігався лише у випадку з 1-м ФВР, таким чином кожен таку серію досліджували, чи це було суттєво чи ні, за допомогою рівняння № 6 і 7, згаданих в методах (Додаткова таблиця: Таблиця 3). Тут перші ФВР H різних ФВР Ідентифікація пацієнта ФВР 1 ФВР 2 ФВР 3 ФВР 4 ФВР 5 ФВР 6 ФВР 7 ФВР 8 ФВР 9 ФВР 10 ФВР 11 ФВР 12 19090 рік (Звичайні дані) 0,8766 1,0164 1,0449 1,0348 1,0192 1,0180 1,0150 1,0160 1,0163 1,0168 1,0179 1,0181 100 (Дані про захворювання) 0,9630 1,0368 1,0223 1,0144 1,0117 1,0117 1,0122 1,0134 1,0142 1,0143 1,0145 1,0148 15 що відповідає пацієнтам (посвідчення особи) 116, 122, 210, 219 та 223 (хворих пацієнтів) спостерігається як незначний і виключається з подальшого аналізу. Отже, аналіз був виконано з 59 серіями даних. Розглядаються H усіх 1-х ФВР усіх часових рядів тут і перелічені в таблиці III. Значення p , перераховане двостороннім t -критерієм студента між H усіх 1 st ФВР, що відповідає даним

про хворобу та нормальним даним, становить $0,3 \times 10^{-9}$ ($p < 0,001$), що вказує на те, що різниця в H , що стосується окремих груп, є значною. У таблиці 3 помітно, що H усіх 1-го ФВР часового ряду здорових пацієнтів варіюється від 0,7886 до 0,9613 із середнім значенням 0,8728 та медіаною 0,8816, тоді як те саме для хворих пацієнтів коливається в межах від 0,8685 до 1,0324 із середнім значенням 0,9940 та медіана 1,0018. Якщо ми нехтуємо значеннями H для пацієнтів (ID) 19093, 19140 та 232 в нашому аналізі, виникнуть зміни з 0,7886 до 0,9297 (із середнім значенням 0,8631 та медіаною 0,8705) для нормальних пацієнтів та 0,9361-1,0324 (із середнім значенням 0,9972 та медіаною 1,0025) для хворих пацієнтів. Отже, існує чітка межа між значеннями H нормального та хворих пацієнтів (рис. 7). Це означає, що для звичайних пацієнтів H завжди менше 0,9300 і це більше, ніж те саме, що відповідає хворим пацієнтам, за винятком вищезазначених трьох пацієнтів. Факт передбачає, що динамічна поведінка серцевої системи в нормі людина помітно відрізняється від тієї ж людини, яка страждає аритмією. Тому твердження відповідає дійсності для даних 56-часового ряду, крім 03, тобто існує ймовірність 94,92% для передбачуваності хвороби аритмії у пацієнта за допомогою цієї процедури аналізу.

Patient ID	H of different IMFs											
	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	IMF 7	IMF 8	IMF 9	IMF 10	IMF 11	IMF 12
19090 (Normal data)	0.8766	1.0164	1.0449	1.0348	1.0192	1.0180	1.0150	1.0160	1.0163	1.0168	1.0179	1.0181
100 (Disease data)	0.9630	1.0368	1.0223	1.0144	1.0117	1.0117	1.0122	1.0134	1.0142	1.0143	1.0145	1.0148

Таблиця 2 - Порівняння H різних ФВР для нормальних даних та даних про хворобу

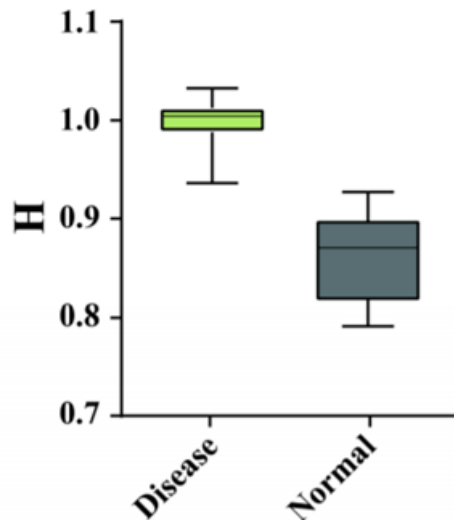


Рисунок 7 - Значення для Н усіх 1-х ФВР хворих та нормальних пацієнтів

Normal data											
Patient ID	19090	19093	19140	19830	16265	16272	16273	16420	16483	16539	16773
H value	0.8766	0.8436	0.9022	0.8644	0.8876	0.8605	0.8002	0.8073	0.9138	0.8261	0.8866
Patient ID	16786	16795	17052	17453	18177	18184	19088				
H value	0.8126	0.7886	0.9297	0.9185	0.9399	0.9613	0.8915				
Diseased data											
Patient ID	100	101	103	105	106	107	108	109	111	112	113
H value	0.9630	1.0018	1.0072	1.0129	0.9966	1.0303	0.9779	1.0115	1.0137	1.0189	1.0077
Patient ID	114	115	117	118	119	121	123	124	200	201	202
H value	1.0044	1.0033	0.9964	1.0202	1.0161	1.0124	0.9869	1.0157	0.9975	0.9965	1.0096
Patient ID	203	205	207	208	209	212	213	214	215	217	220
H value	0.9950	1.0061	0.9405	0.9994	0.9544	0.9821	0.9714	1.0203	0.9361	1.0324	0.9847
Patient ID	221	222	228	230	231	232	233	234			
H value	1.0004	0.9792	0.9687	0.9885	1.0115	0.8685	1.0110	1.0042			

Таблиця 3 - Значення Н перших ФВР даних кожного пацієнта (здорового та з аритмією)

Моменти продукту Пірсона (r) між значеннями Н першого ФВР (для хворих, а також здорових піддослідних) і відповідний вік осіб був розрахований для перевірки лінійної залежності між вищезазначеними змінними. Значення r

для нормальної пацієнтів оцінювали як позитивні (0,2671), тоді як згадане значення було негативним (-0,2645) для хворих пацієнтів. Отже, позитивна (до певної міри) кореляція присутня для здорових піддослідних та негативна (до певної міри) кореляція відображається для хворих пацієнтів. Графіки розсіювання разом з еліпсом впевненості для нормальних випадків та випадків захворювання ілюструються на рис. 8. Отже, результат також підсилює згаданий висновок. Дані про хворобу та нормальні дані пацієнта також можна розрізнити за значеннями «r».

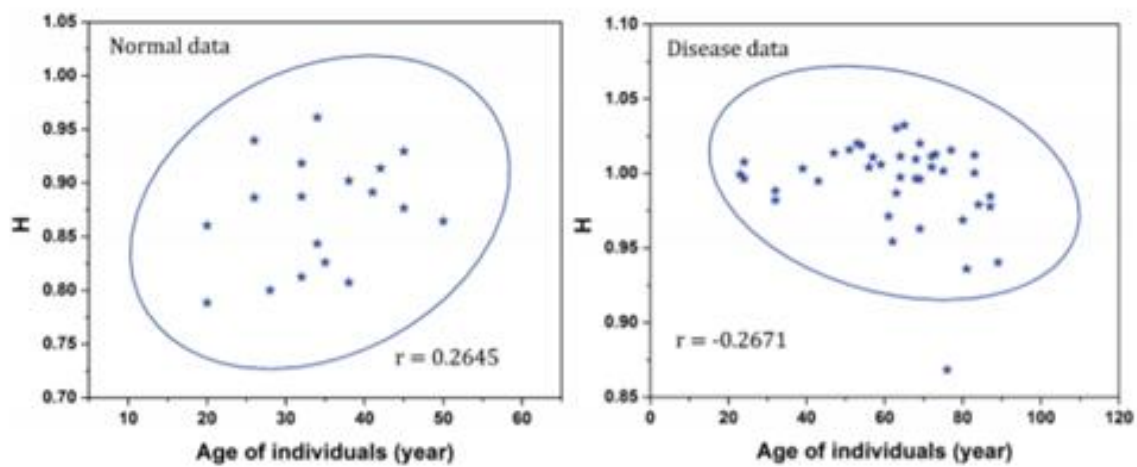


Рисунок 8 - Розсіяний графік разом з еліпсом впевненості для H усіх 1-х ФВР нормальних та хворих пацієнтів

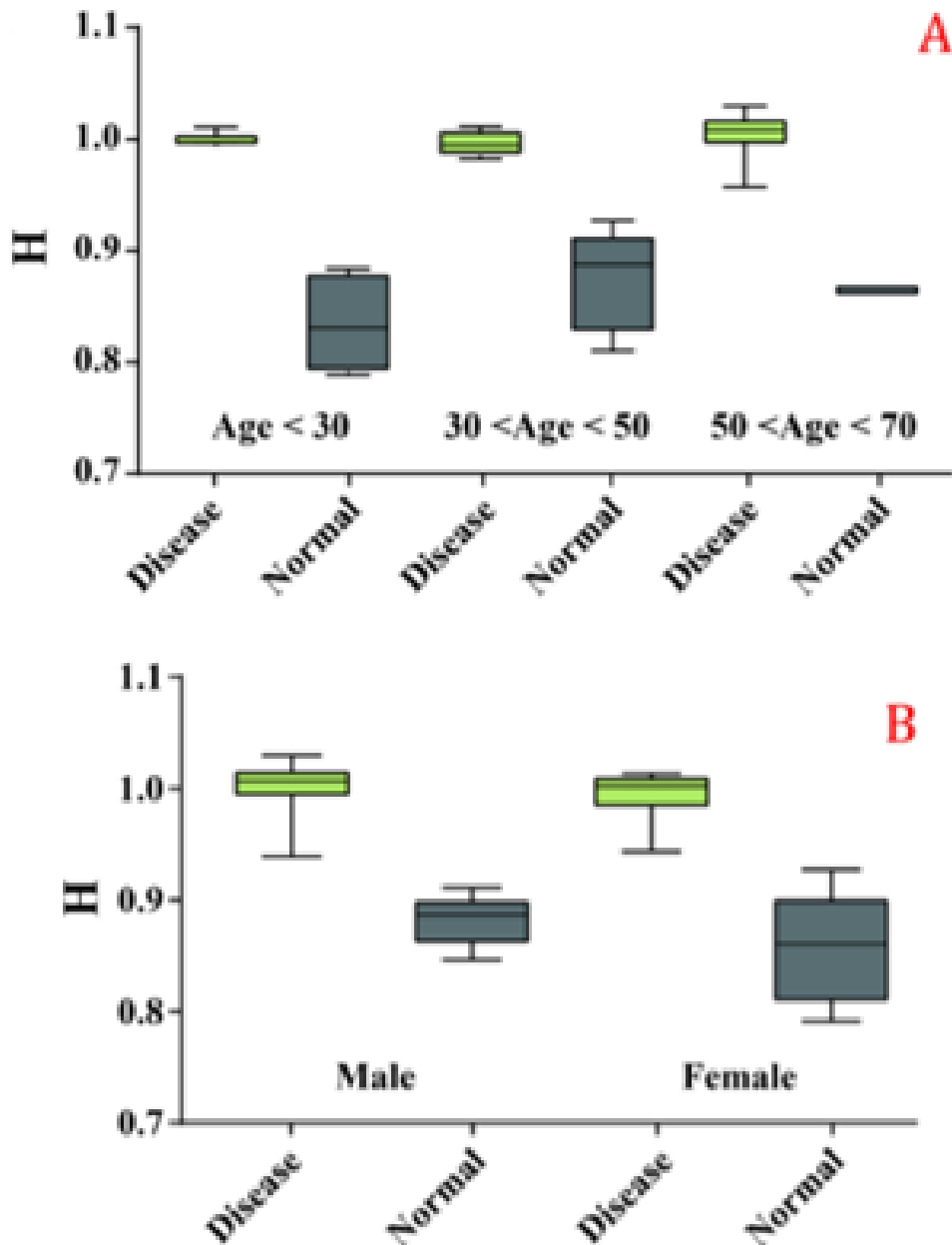


Рисунок 9 - Графік значень для Н усіх 1-х ФВР нормальних та хворих пацієнтів відповідно

до вікової групи(зверху) та гендерної групи(знизу)

2.2 Розгляд результатів за різними статево-віковими групами

Результат роботи також може бути підтверджений посиланням на дані пацієнтів за гендерним та віковим угруповання. Тут такий розподіл призведе до того, що ми оцінимо, чи Н (тільки перша. ФВР) розподіляється рівномірно у всьому діапазоні, а також підкоряється подібній схемі на короткому

діапазоні класифікації. У зв'язку з цим H усіх 1-го ФВР хворих та нормальних пацієнтів класифікують за їх статтю, і те саме подано в таблиці 4. Середнє значення H хворих пацієнтів чоловічої статі становить 0,9992 (медіана = 1,0061), тоді як те ж саме для норми у пацієнтів чоловічої статі 0,8816 (медіана = 0,8866). 0,9950 (1 0006) та 0,8547 (0,8644) є середнім значенням (медіана) H , що відповідає хворим та нормальним пацієнтам жінкам відповідно. Дискретна межа між H нормальних та хворих пацієнтів також узгоджується з гендерне групування (рис. 9). З іншого боку, таблиця 5 відображає деталі H усіх 1 вул ФВР на основі вікової групи пацієнтів. Значення H у пацієнтів віком до 30 років що відповідає захворюванню, а нормальні суб'єкти мають середні значення 1,0013 (медіана = 0,9994) та 0,8552 (медіана = 0,8736) відповідно. Середні (медіанні) значення H захворювання та нормальні випадки для пацієнтів у віці від 30 до 50 років оцінюються як 0,9965 (0,9950) та 0,8809 (0,8821) послідовно. Те саме для пацієнтів з віком від 18 (Таблиця 4): Статистичні показники H 1-го ФВР даних нормальних та хворих пацієнтів за гендерною групою Від 50 до 70 відображаються у зазначеній таблиці як 1,0026 (медіана = 1,0078) та 0,8644 відповідно (Рис. 9). Однак оцінка на тій же платформі для пацієнтів у віці (> 70 років) є пропущено через відсутність відповідних даних нормальних пацієнтів для цієї вікової межі.

	Patient ID	Age	H of 1st IMF	Mean	Median	p-value	Patient ID	Age	H of 1st IMF	Mean	Median	p-value
	Male patients						Female patients					
Disease data	230	32	0.9885	0.9992	1.0061	0.0001	208	23	0.9994	0.9950	1.0006	0.000002
	203	43	0.9950				106	24	0.9966			
	214	53	1.0203				113	24	1.0077			
	112	54	1.0189				212	32	0.9821			
	233	57	1.0110				115	39	1.0033			
	205	59	1.0061				111	47	1.0137			
	213	61	0.9714				119	51	1.0161			
	209	62	0.9544				234	56	1.0042			
	107	63	1.0303				123	63	0.9869			
	109	64	1.0115				114	72	1.0044			
	200	64	0.9975				231	72	1.0115			
	217	65	1.0324				105	73	1.0129			
	201	68	0.9965				101	75	1.0018			
	202	68	1.0096				228	80	0.9687			
	100	69	0.9630				121	83	1.0124			
	117	69	0.9964				222	84	0.9792			
	118	69	1.0202				108	87	0.9779			
	124	77	1.0157				220	87	0.9847			
	215	81	0.9361				207	89	0.9405			
	221	83	1.0004									
103	NR ^a	1.0072										
Normal Data	16773	26	0.8866	0.8816	0.8866	0.8547	16272	20	0.8605	0.8644		
	16265	32	0.8876				16795	20	0.7886			
	19093	34	0.8436				16273	28	0.8002			
	16483	42	0.9138				16786	32	0.8126			
	19090	45	0.8766				17453	32	0.9185			
							16539	35	0.8261			
							19140	38	0.9022			
							16420	38	0.8073			
							19088	41	0.8915			
							17052	45	0.9297			
							19830	50	0.8644			

Таблиця 4 - Статистичні показники Н 1-го ФВР даних нормальних та хворих пацієнтів за гендерними групами

Age group	Patient ID	Age (M/F)	H of 1st IMF	Mean	Median	Patient ID	Age (M/F)	H of 1st IMF	Mean	Median	p-value
Age < 30	208	23 (F)	0.9994	1.0013	0.9994	16272	20 (F)	0.8605	0.8552	0.8736	0.005215
	106	24 (F)	0.9966			16795	20 (F)	0.7886			
	113	24 (F)	1.0077			16773	26 (M)	0.8866			
30 < Age < 50	212	32 (F)	0.9821	0.9965	0.9950	16273	28 (F)	0.8002	0.8809	0.88209868	0.000001
	230	32 (M)	0.9885			16265	32 (M)	0.8876			
	115	39 (F)	1.0033			16786	32 (F)	0.8126			
	203	43 (M)	0.9950			17453	32 (F)	0.9185			
	111	47 (F)	1.0137			19093	34 (M)	0.8436			
						16539	35 (F)	0.8261			
						19140	38 (F)	0.9022			
						16420	38 (F)	0.8073			
						19088	41 (F)	0.8915			
						16483	42 (M)	0.9138			
						19090	45 (M)	0.8766			
						17052	45 (F)	0.9297			
50 < Age < 70	119	51 (F)	1.0161	1.0026	1.0078	19830	50 (F)	0.8644	0.8644	0.8644	
	214	53 (M)	1.0203								
	112	54 (M)	1.0189								
	234	56 (F)	1.0042								
	233	57 (M)	1.0110								
	205	59 (M)	1.0061								
	213	61 (M)	0.9714								
	209	62 (M)	0.9544								
	107	63 (M)	1.0303								
	123	63 (F)	0.9869								
	109	64 (M)	1.0115								
	200	64(M)	0.9975								
	217	65 (M)	1.0324								
	201	68 (M)	0.9965								
	202	68 (M)	1.0096								
	100	69 M)	0.9630								
	117	69 (M)	0.9964								
118	69 (M)	1.0202									

Таблиця 5 - Статистичні показники H 1-го ФВР даних нормальних та хворих пацієнтів відповідно до вікової групи

у цій таблиці розрізняють H , що відповідає хворим та нормальним також зустрічається незалежно від оцінки, яка проводиться відповідно до віку пацієнтів. Ось, значення p -тесту для будь-якого групування також перелічені у вищезазначених таблицях та вони показують результати, які будуть прийняті до цього. Усі 1-й ФВР потрапляють до категорії поведінка типу персистенції як H однакових завжди перевищує 0,5. Однак усі H для дані хвороби (~ 1) перевищують дані звичайних даних. Справа в тому, що 1-й ФВР 20 Серія даних про захворювання демонструє більш сильну тенденцію, ніж у звичайному випадку.

РОЗДІЛ 3

РЕАЛІЗАЦІЯ. ІНСТРУМЕНТИ ТА ЗАСОБИ

3.1 Інструменти та засоби

Інструментом для розробки системи була мова програмування Python версії 3.6.3. Вибір був обумовлений наявністю широкого асортименту бібліотек для різного роду математичних обчислень, обробки, тощо.

Для дослідження були взяті записи ЕКГ однакової довжини (1 хв) з одним відведенням з ресурсу physionet.org. Дані знаходяться у форматі CSV та мають 2 поля: час, дата та значення сигналу в мікрвольтах (рис. 10). У мові Python доступна зручна відкрита бібліотека Pandas для обробки даних в тому числі й такого формату. Її ми і застосували для читання та попередньої обробки.

```
'Time and date','ECG '  
'hh:mm:ss.mmm dd/mm/yyyy','mV'  
'[01:11:04.000 01/11/2000]',0.567  
'[01:11:04.003 01/11/2000]',0.673  
'[01:11:04.007 01/11/2000]',0.774  
'[01:11:04.010 01/11/2000]',0.867  
'[01:11:04.013 01/11/2000]',0.947  
'[01:11:04.017 01/11/2000]',0.985  
'[01:11:04.020 01/11/2000]',0.977  
'[01:11:04.023 01/11/2000]',0.945  
'[01:11:04.027 01/11/2000]',0.930  
'[01:11:04.030 01/11/2000]',0.919  
'[01:11:04.033 01/11/2000]',0.912  
'[01:11:04.037 01/11/2000]',0.906
```

Рисунок 10

Для візуалізації була використана також відкрита бібліотека під назвою Matplotlib (рис. 11).

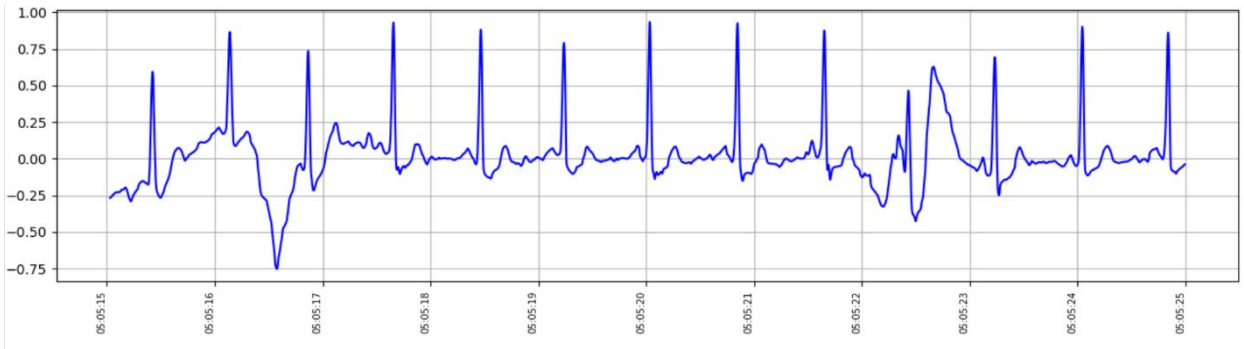


Рисунок 11

Для пошуку деяких параметрів сигналу була використана бібліотека Scipy.

Щоб знайти частоту серцевих скорочень необхідно було знайти так звані «R-піки» (рис. 12).

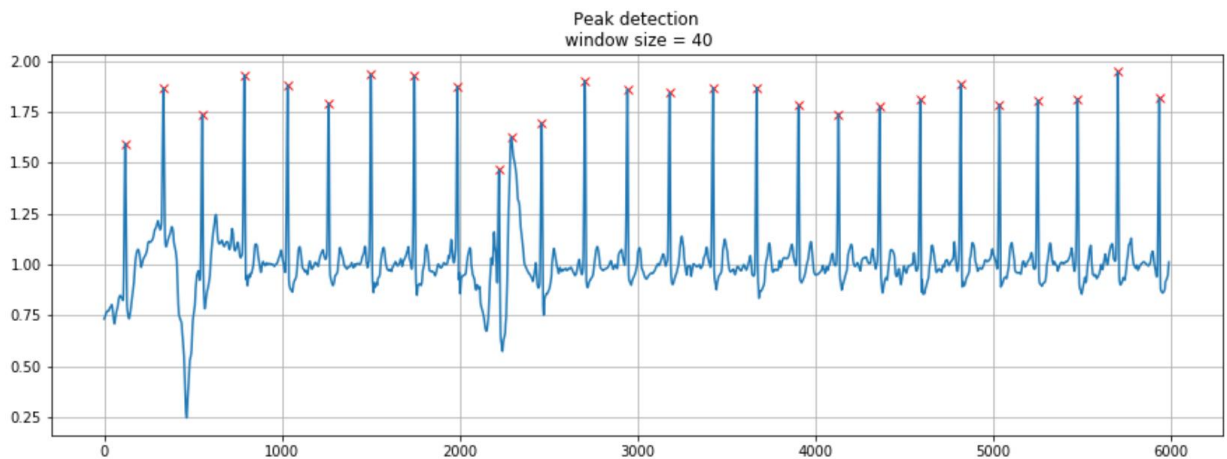


Рисунок 12

3.2 Реалізація мовою програмування Python

Оглянемо реалізацію фільтру Савіцького-Голея:

```
from scipy import signal
from scipy.signal import butter, iirnotch, lfilter
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
## A high pass filter allows frequencies higher than a cut-off value
def butter_highpass(cutoff, fs, order=5):
    nyq = 0.5*fs
    normal_cutoff = cutoff/nyq
    b, a = butter(order, normal_cutoff, btype='high', analog=False, output='ba')
    return b, a
## A low pass filter allows frequencies lower than a cut-off value
def butter_lowpass(cutoff, fs, order=5):
    nyq = 0.5*fs
    normal_cutoff = cutoff/nyq
    b, a = butter(order, normal_cutoff, btype='low', analog=False, output='ba')
    return b, a
def notch_filter(cutoff, q):
    nyq = 0.5*fs
    freq = cutoff/nyq
    b, a = iirnotch(freq, q)
    return b, a

def highpass(data, fs, order=5):
    b, a = butter_highpass(cutoff_high, fs, order=order)
    x = lfilter(b, a, data)
    return x
```

За допомогою коду, наявного вище прибираємо зайві шуми з даних ЕКГ

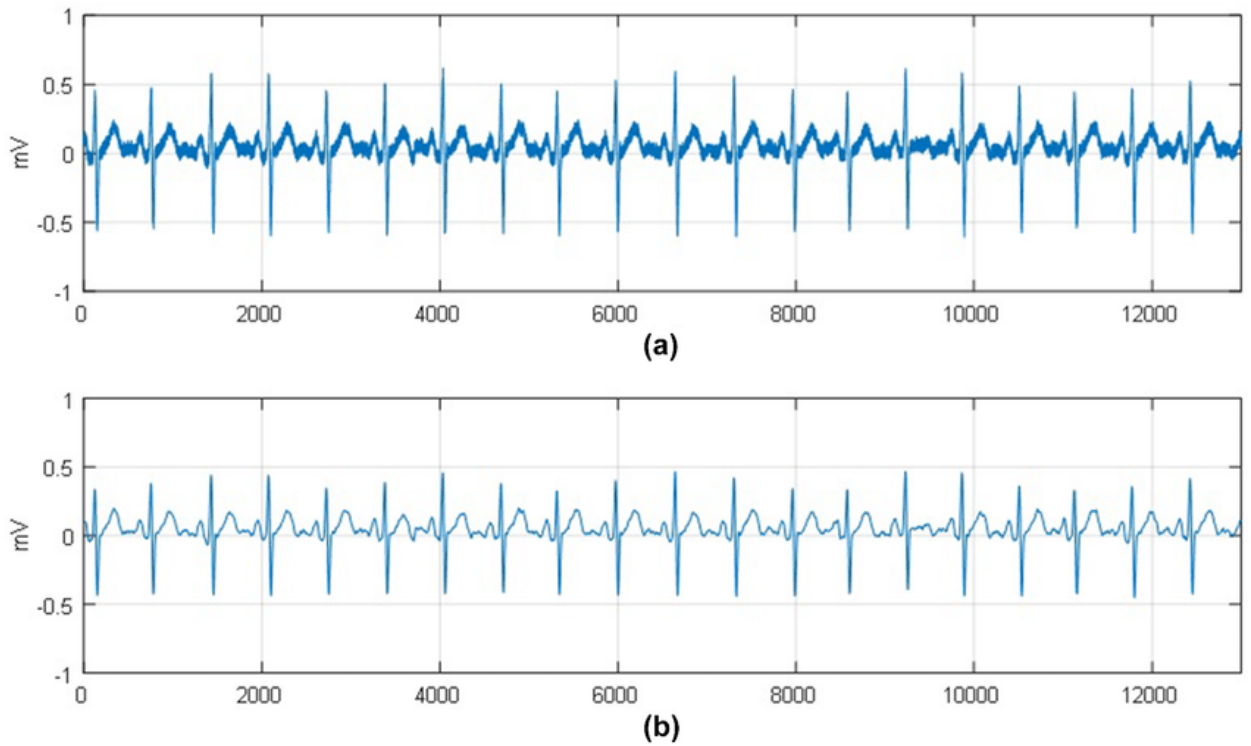


Рисунок 13 – Приклад роботи з видалення шумів фільтром Савіцького-Голя

```
def lowpass(data, fs, order =5):
    b,a = butter_lowpass(cutoff_low, fs, order=order)
    y = lfilter(b,a,data)
    return y

def notch(data, powerline, q):
    b,a = notch_filter(powerline,q)
    z = lfilter(b,a,data)
    return z

def final_filter(data, fs, order=5):
    b, a = butter_highpass(cutoff_high, fs, order=order)
    x = lfilter(b, a, data)
    d, c = butter_lowpass(cutoff_low, fs, order = order)
    y = lfilter(d, c, x)
    f, e = notch_filter(powerline, 30)
    z = lfilter(f, e, y)
    return z

ecg_signal = np.loadtxt('df_ecg_Polar.csv', skiprows=0)
fs = 1000
## Order of five works well with ECG signals
cutoff_high = 0.5
cutoff_low = 2
powerline = 60
order = 5
```

```
plt.figure(1)
ax1 = plt.subplot(211)
plt.plot(ecg_signal)
ax1.set_title("Raw ECG signal")

filter_signal = final_filter(ecg_signal, fs, order)
ax2 = plt.subplot(212)
plt.plot(filter_signal)
ax2.set_title("Clean ECG signal")
plt.show()
```

```

try:
    window_size = np.abs(np.int(window_size))
    order = np.abs(np.int(order))
except ValueError, msg:
    raise ValueError("window_size and order have to be of type int")
if window_size % 2 != 1 or window_size < 1:
    raise TypeError("window_size size must be a positive odd number")
if window_size < order + 2:
    raise TypeError("window_size is too small for the polynomials order")
order_range = range(order+1)
half_window = (window_size - 1) // 2
# precompute coefficients
b = np.mat([[k**i for i in order_range] for k in range(-half_window, half_window+1)])
m = np.linalg.pinv(b).A[deriv] * rate**deriv * factorial(deriv)
# pad the signal at the extremes with
# values taken from the signal itself
firstvals = y[0] - np.abs( y[1:half_window+1][::-1] - y[0] )
lastvals = y[-1] + np.abs(y[-half_window-1:-1][::-1] - y[-1])
y = np.concatenate((firstvals, y, lastvals))
return np.convolve( m[::-1], y, mode='valid')

```

```

import numpy as np
frequency = 1
total_time = 1
res = 0.01
time = np.arange(0, total_time, res)
amplitude = np.sin(2 * np.pi * frequency * total_time * (time / time[-1]))
np.mean(amplitude)
-2.6319088627735572e-17 # this value literally equal 0

```

За допомогою програми вираховуємо дані, які характеризують сигнал ЕКГ

```

C:/1.py
H of 1st IMF = 0.95627
Disease data

Process finished with exit code 0

```

. ВИСНОВКИ

У цій роботі методика ЕРР обробляє нелінійні та нестационарні ЕКГ-сигнали для діагностики серцево-судинних захворювань. ФВР, який є фундаментально розкладним сигналом цілого ЕКГ-сигналу, що був обробленим для хворих з аритмією пацієнтів, а також для звичайних піддослідних, за допомогою техніки ЕРР. Порівняльне дослідження на основі перерахованих Н ФВР проводиться для розрізнення аритмії у людей. У цьому процесі аналізується чітко визначена межа між Н нормальної та хворої людини, що вказує на контраст у динаміці серцевої діяльності між нормальними та хворими особами. Дослідження показує, що точність цього методу становить 94,92%. Однак це аналіз обмежених конкретних наборів даних про хворих. Можна проводити систематичний аналіз на основі інших даних серцево-судинних захворювань, щоб забезпечити пропорційність прогнозованості. Більше того, рішення може бути перевірено за допомогою застосування інших різних статистичних методів, таких як вейвлет-аналіз, багатофункціональний аналіз та багатофункціональний багатодіагностичний аналіз іншого Аналіз, що поєднується з машинним навчанням, може спричинити трансформацію в поточному сценарій надання медичних послуг. Враховуючи отримані результати та ефективність методу було створено програмний засіб мовою Python, як найвдалішої мови для роботи зі складними науковими алгоритмами, було використано чималу кількість бібліотек, які прискорили роботу і написання програми

Загалом, за підсумками роботи були вирішені такі задачі:

1. Були досліджено статистичні методи аналізу ЕКГ.
2. Вироблено алгоритм обробки ЕКГ.
3. Перевірено ефективність роботи алгоритму на даних пацієнтів.
4. Розроблено програмний засіб для автоматизації роботи алгоритму

В майбутньому є плани повернутися до роботи з цим питання, адже тема клінічної діагностики є надзвичайно актуальною. Також, її вага збільшуватиметься також за рахунок поліпшення апаратної частини. Таким чином, використання подібних до даної систем стане більше широким серед усіх верств населення.

Одним з наступних кроків є покращення ефективних методів обробки зашумлених сигналів ЕКГ та дослідження методів класифікації для діагностики більш широкого спектру відхилень серцево-судинної системи.

Загалом, одним з ключових питань є створення користувацького інтерфейсу, задля спрощення роботи такого та подібних інструментів в стінах медичних закладів, створення кросплатформеного додатку та загалом розширення функціоналу системи.

З розвитком даного дослідження існують перспективи впровадження даної системи до процесу клінічної діагностики медиками в якості допоміжного засобу для спрощення інтерпретації ЕКГ. В ситуаціях, де має місце дефіцит кваліфікованих медиків, така система буде особливо актуальною і зможе збільшити ефективність праці медичного персоналу. Також, можливе впровадження в якості програмного забезпечення для побутової електроніки, як наприклад фітнес браслети з функцією ЕКГ. Існують люди, що входять до групи ризику певних серцево-судинних захворювань. В такому випадку їм вкрай важливо мати доступний спосіб самостійного моніторингу для вчасного виявлення та запобігання захворювань та відхилень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Agarwal, S., Rani, A., Singh, V., & Mittal, A. P. (2016). Performance Evaluation and Implementation of FPGA Based SGSF in Smart Diagnostic Applications. *Journal of Medical Systems*, 40(3), 1–15. <https://doi.org/10.1007/s10916-015-0404-2>
2. Alwan et al. (2010). *Global status report on noncommunicable diseases 2010*. World Health Organization Press, Geneva, Switzerland. <https://doi.org/10.1017/CBO9781107415324.004>
3. Barman, C., Ghose, D., Sinha, B., & Deb, A. (2016). Detection of earthquake induced radon precursors by Hilbert Huang Transform. *Journal of Applied Geophysics*, 133, 123–131. <https://doi.org/10.1016/j.jappgeo.2016.08.004>
4. Berkaya, S. K., Uysal, A. K., Sora Gunal, E., Ergin, S., Gunal, S., & Gulmezoglu, M. B. (2018). A survey on ECG analysis. *Biomedical Signal Processing and Control*, 43, 216–235. <https://doi.org/10.1016/j.bspc.2018.03.003>
5. Blanco-Velasco, M., Weng, B., & Barner, K. E. (2008). ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology and Medicine*, 38(1), 1–13. <https://doi.org/10.1016/j.compbiomed.2007.06.003>
6. Chatfield, C. (2004). *The analysis of time series—an introduction*. Chapman and Hall, London.
7. Chaudhuri, H., Seal, K., Maji, C., Pal, S. & Mandal, M. K. (2019). The unrevealed facts on helium resources of India. *Arabian Journal of Geosciences*, 12, 216 (1-19). <https://doi.org/10.1007/s12517-019-4369-1>
8. Dahiru, T. (2008). P-Value, a true test of statistical significance? a cautionary note. *Annals of Ibadan Postgraduate Medicine*, 6(1), 21–26. <https://doi.org/10.4314/ajpm.v6i1.64038>
9. Davis, N. V. (2012). Feature Extraction Using Empirical Mode Decomposition of Speech Signal. 3, 77–80.
10. de Godoy, M. F. (2016). Nonlinear Analysis of Heart Rate Variability: A Comprehensive Review. *Journal of Cardiology and Therapy*, 3(3), 528–533. <https://doi.org/10.17554/j.issn.2309-6861.2016.03.101-4>
11. Fell, J., Mann, K., Röschke, J., & Gopinathan, M. S. (2000). Nonlinear analysis of continuous ECG during sleep II. Dynamical measures. *Biological Cybernetics*, 82(6), 485–491. <https://doi.org/10.1007/s004220050601>

12. Feng, C., & Vidakovic, B. (2015). Estimation of the Hurst Exponent Using Trimean Estimators on Nondecimated Wavelet Coefficients. *Journal of Latex Class Files* 14(8), 1–11. Retrieved from <http://arxiv.org/abs/1709.08775>
13. Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng C-K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, 101(23). <https://doi.org/10.1161/01.cir.101.23.e215>
14. Hansson, A., Madsen-Härdig, B., & Olsson, S. B. (2004). Arrhythmia-provoking factors and symptoms at the onset of paroxysmal atrial fibrillation: A study based on interviews with 100 patients seeking hospital assistance. *BMC Cardiovascular Disorders*, 4, 1–9. <https://doi.org/10.1186/1471-2261-4-13>
15. Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116, 770-799
16. John, R. M., & Kumar, S. (2016). Sinus node and atrial arrhythmias. *Contemporary Reviews in Cardiovascular Medicine*, 133, 1892–1900. <https://doi.org/10.1161/CIRCULATIONAHA.116.018011>
17. Mandelbrot, B. B., & Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, 5(5), 967-988. <https://doi.org/10.1029/WR005i005p00967>
18. Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3), 45–50. <https://doi.org/10.1109/51.932724>
19. Savitzky, A., & Golay M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627-1639
20. Subramanian, B. (2017). ECG signal classification and parameter estimation using multiwavelet transform. *Biomedical Research (India)*, 28(7), 3187–3193.
21. PhysioNet: <https://physionet.org/>